DOCUMENT RESUME

ED 041 043 24 TM 000 008

AUTHOR Jones, Margaret Hubbard
TITLE The Unintentional Memory Load in Tests for Young Children.
INSTITUTION California Univ., Los Angeles. Center for the Study of Evaluation.
SPONS AGENCY Research and Development Centers Branch, DEL.
REPORT NO CSE-57
BUREAU NO BR-6-1646
PUB DATE May 70
CONTRACT OEC-4-6-06146-1909
NOTE 27p.

EDRS PRICE EDRS Price MF-$0.25 HC-$1.45
DESCRIPTORS *Comprehension, Diagnostic Tests, *Learning Processes, Listening Comprehension, *Memory, *Primary Education, Reading Comprehension, Retention, *Standardized Tests, Test Construction, Tests, Test Validity

ABSTRACT

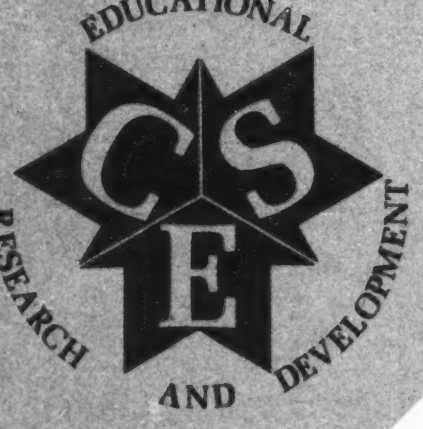
The validity of certain standardized tests may be affected by the short-term memory load therein and its relation to a child's short-term memory capacity. Factors of testing which increase a test's memory load and consequently interfere with comprehension are discussed. It is hypothesized that a test which strains the short-term memory capacity of a child is functioning as an unintended screen to eliminate subjects with inferior memory capacities and does not adequately measure the intended ability. A number of tests randomly selected are examined in terms of three ways in which memory capacities can be overloaded in a test. In conclusion, since there does not appear to be a high correlation between memory skills and the other skills involved in the learning processes, it is recommended that memory skills be measured as such and that the other skills be assessed independently of the memory screen. A listing of tests is included. (PR)

ED041043

BRG-1646
PA24
TM

CENTER FOR THE
STUDY OF
EVALUATION

UCLA
Graduate School
of Education
Los Angeles, California



THE UNINTENTIONAL MEMORY LOAD IN TESTS FOR YOUNG CHILDREN

Margaret Hubbard Jones

CSE Report No. 57
May 1970

TM 000 008

CENTER FOR THE STUDY OF EVALUATION

Marvin C. Alkin, Director
Publications Committee:
James W. Trent, Chairman
Theodore R. Husek
Sherman J. Pearl
Audrey Schwartz

UCLA Graduate School of Education

The CENTER FOR THE STUDY OF EVALUATION is one of nine centers for educational research and development sponsored by the United States Department of Health, Education and Welfare, Office of Education. The research and development reported herein was performed pursuant to a contract with the U.S.O.E. under the provisions of the Cooperative Research Program.

Established at UCLA in June, 1966, CSE is devoted exclusively to finding new theories and methods of analyzing educational systems and programs and gauging their effects.

The Center serves its unique functions with an inter-disciplinary staff whose specialties combine for a broad, versatile approach to the complex problems of evaluation. Study projects are conducted in three major program areas: Evaluation of Instructional Programs, Evaluation of Educational Systems, and Evaluation Theory and Methodology.

This publication is one of many produced by the Center toward its goals. Information on CSE and its publications may be obtained by writing:

Office of Dissemination
Center for the Study of Evaluation
UCLA Graduate School of Education
Los Angeles, California 90024

ED041043

THE UNINTENTIONAL MEMORY LOAD IN TESTS FOR YOUNG CHILDREN

By

Margaret Hubbard Jones

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

CSE Report No. 57

May 1970

Center for the Study of Evaluation
UCLA Graduate School of Education
Los Angeles, California

THE UNINTENTIONAL MEMORY LOAD IN TESTS FOR YOUNG CHILDREN

Although learning has long been considered the crux of the education process and much research has been devoted to it, little cognizance has been taken of memory by educators,[1] except to advocate drill where certain skills obviously must be mastered, or to deprecate dependence upon "memory" where more complex cognitive processes are thought desirable. Memory processes are essential to any learning and to any comprehension of language, yet they seem to be so poorly understood that their characteristics are not used to advantage and their presence is apparently not recognized. This state of affairs cannot be attributed to lack of basic research, for the literature in the field of memory has become vast in the last ten years, and the characteristics of at least two processes, Short-term Memory (STM) and Long-term Memory (LTM) have been rather widely explored.[2] The difficulty probably lies in the usual gap between "pure" research and its application in a more complex situation. It is with the hope that this gap can be bridged that this paper is undertaken.

---

[1] Neither The Handbook of Research on Teaching (N.L. Gage, Ed., 1963) nor Tyler's Psychology of Human Differences (1965) treats memory.

[2] For a very brief review at the elementary level, cf. Norman, 1969. As Norman says: "All paths lead to memory." p. 177.

There are, of course, many areas in education where a survey of the contribution of various memory processes would be valuable, but there are two areas in which an indication of the extent of such contribution is relatively more important, both because it is in these cases a very powerful limiting factor and because it appears to be unsuspected. These areas are initial reading instruction and standardized tests. The first of these is discussed elsewhere (Jones, 1970). Here we will consider the short-term memory load in a number of typical standardized tests of several sorts as it relates to the short-term memory capacity of young children, and the implications that these findings have for the validity of the tests.

Short-term memory is a specific kind of ability or skill[3] showing wide individual differences (e.g., Guilford, 1967; Jones, 1970; Hurlock & Newmark, 1931). It is limited in capacity, minimally coded, lasts (in adults) a maximum of thirty seconds and degrades rapidly after about four seconds. It is the kind of memory process used for retaining unfamiliar telephone numbers long enough to dial them. It is the "magic number 7 $\pm$ 2"--except that this is a gross overestimate of normal adults' STM capacity, which is closer to 5 $\pm$ 2 (e.g., Atkinson & Shiffrin, 1967) or even 4 $\pm$ 2 (Ross, 1969; Mortenson & Loess, 1964). It is one of the essential parts of the information processing mechanism. It is the critical process in comprehension of language: All parts of a sentence

---

[3] Or a group of skills (Guilford, 1967; Bradley, Hoepfner & Guilford, 1969).

must be retained in STM until the string is concluded and comprehended; a sentence cannot be comprehended piecemeal because the particular meaning of the words is only clear as their sequence and the syntax reveals it.[4]

The capacity of STM grows with maturity and also with training. There are likewise wide individual differences in capacity at all ages and educational levels. Young children have a very small capacity--even the average second grader cannot easily handle more than three "units," and the capacity of the store grows slowly. One of the problems which occurs in attempting to match learning or testing materials to a child's STM capacity is the determination of what constitutes a unit: "Words" are only occasionally units; the size of the units varies with experience, verbal ability, and maturity (cf. Jones, 1970, for a fuller discussion; Mathews, 1968; and Broadbent, 1965). Nevertheless, a conservative estimate of the lower bound for the number of units involved in a test item or in test instructions can often be made by counting the number of semantic items that must be stored in memory to enable the student to answer the specific question asked. Such an estimate will err on the low side, because it assumes that the student comprehends each clause and stores it as a single semantic unit in memory. He may

---

[4] Many "words," particularly those most frequently used in English, have multiple meanings. Quite obviously, it is the context which permits the listener (or reader) to assign the proper meaning to the word symbols. Even with simple words, a precise reading cannot be given without syntactic context, e.g., "bowl" or "hit" or "call" may be either nouns or verbs.

not be able to do this with all the strings for one reason or another. If he cannot, then he must remember more items--some of the individual words or phrases verbatim--which will always add to the total memory load.

Some of the factors which interfere with comprehension (often quite unnecessarily) by increasing the memory load are: (1) difficult or unusual syntax (for example; conditional statements, word order which does not match the order of events, or deletion tranforms, cf., Kennedy, 1970 a,b); (2) difficult lexical items ("derelict," "dinghy" in tests of comprehension, not of vocabulary, at third grade, or "enthusiastic" as a response choice in a first grade vocabulary test, or "engage in" as part of a sentence for defining a first grade vocabulary item (vide infra); (3) ambiguous statements;[5] (4) interference, in the form of interpolated (often unnecessary) instruction (cf. Mortenson & Loess, 1964); (5) delay (e.g., Calfee, 1969; Whimbey & Leiblum, 1967), again in the form of interpolated instructions or, more often, mere excess verbiage that adds nothing; (6) rate of presentation (Sitterly, 1968, Hansen, 1965, Gordon, Gordon & Perrier, 1967, Schulman & Lovelace, 1967);

---

[5] There is no research on this point, but probably ambiguity is not the same for adults and for children. Adults with high verbal ability know many of the multiple meanings of common English words; children, even highly verbal older children, typically know very few of these, even though they can give a "definition" of the word satisfactory to teachers or test constructors. Hence children may not see ambiguities that adults see. On the other hand, partly because of weaker contextual constraints on perception, children may mishear words more frequently than adults, and find ambiguities where adults do not suspect them. This important pedagogical problem deserves some research.

(7) ambient noise (Aaronson & Sternberg, 1964). If any of these factors are introduced into a test (either achievement or aptitude), the short-term memory load will be increased above the minimal estimate. Furthermore, if the short-term memory capacity is exceeded, then damage may be done to the contents of the store, and less is retained than would normally be the case (Sperling & Speelman, 1967; Mackworth, 1964; Lloyd, 1961). In addition, any work which has to be done on the contents of short-term memory--transformations of any sort, such as converting "dime" to cents or adding two numbers, will require part of the storage capacity (cf., Posner & Rossman, 1965; Talland, 1967; Jones, 1968). In sum, short-term memory capacity is limited; it is easily damaged by a variety of commonly occurring events; it is a necessary prerequisite for comprehension. Therefore, any test which strains the short-term memory capacity of children of the age for which the test was intended will, first and foremost, be a test of memory and only secondarily will it measure what it was intended to measure. That is, short-term memory acts as a screen to eliminate those students with less than superior talent for remembering; for such students the intended ability is never measured at all. There is no reason to believe that short-term memory is closely related to other cognitive skills (cf., Whimbey, Fischhof, & Silikowitz, 1969; Guilford, 1967). There is good reason to believe that not only are memory-for-item and memory-for-sequence of item different skills (Guilford, 1967; Moore & Ross, 1963; Jones, 1970), but that memory for visual items is different from memory for auditory

items (Orpet & Meyers, 1966; Meeker, 1966; Murdock & Walker, 1969; Feldman, 1970). These facts call into serious question the real validity of the tests which utilize, albeit inadvertently, a memory screen, if validity be taken in the old-fashioned sense of testing what it is supposed to test. The practice of reporting as validity a correlation with some test equally ill-defined is hardly comforting.

The short-term memory span of young children is indeed short, a fact which is apparently not common knowledge. Even under ideal conditions (lists of monosyllabic count nouns of very high frequency), third graders have a mean auditory memory span (recall) of less than 4 items (Jones, 1968). Ingles and Sykes (1967) point out that STM increases slowly between five and ten years. With even simpler tests, using non-verbal items and a recognition procedure (pointing as the response), children of nursery school age have been shown to have a capacity of only two items (Calfee, Hetherington, & Waltzer, 1966; cf., also Haith, 1968). Even college students of high verbal ability can reliably retrieve only 5 items (Phillips, Shiffrin, & Atkinson, 1967). This is, then, a very critical limitation for both test instructions and test items.

A number of standardized tests of both achievement and aptitude at the primary level have been examined, with an eye to identifying those containing an unintended memory screen. The tests were selected at random, and those cited are identified only in order to make a general point or to show that the faults are not limited to lesser known tests, but are prominent in some of the most widely used and more prestigious

tests. There are three somewhat different ways in which memory capacity can be overloaded: (1) in the test instructions; (2) within the individual items; or (3) in the temporal relations between the questions and the original presentation of information. Each is an important source of variance in its own right. Of the three, the second is the commonest, probably because it is the least obvious. Much has been written about the necessity for simple instructions, but apparently this caution is often interpreted to mean merely that the lexical difficulty should be low, rather than that, in addition, the procedure should be made sufficiently simple for a child to follow without exceeding his memory capacity. The third kind of memory load occurs relatively infrequently; most test constructors have a general understanding of the difficulties which both delay and interference cause for memory processes. Since the tests examined were largely those intended for the early primary school years, most of them are concerned with language or reading skills, but several "intelligence" tests and arithmetic tests were also included, as well as a test of musical interest, and a picture interest inventory. No attempt at completeness was made because the point to be made is a general one, rather than a specific critique of all tests. The citations should be taken as examples of one of the general difficulties that run throughout the testing of young children.

Test Instructions

Difficult and involved instructions seem particularly out of place in tests for children; it is even more disturbing when they accompany

a test which is otherwise well designed and constructed. An example of such a test is the Gates-MacGinitie Test of Readiness Skills for kindergarten and first grade (Gates-MacGinitie Reading Tests, 1965). The initial instructions to be read to the children for the first subtest: Listening Comprehension, are:

> I'm going to read some short stories. For each story there is a box of three pictures in your booklet. After each story, I'll read a question. You will answer the question by putting an X on one of the three pictures. I'll read each story only once, so you will have to listen carefully. If you make a mistake, you may erase it and mark the right picture. We'll do the first story now.
>
> Take your pencil, ready to work. Put one finger of your other hand on the first box, the box that has the thread, scissors, and pin in it, right here where I'm pointing . . . Listen carefully to the story and the question. Don't make any marks until I stop reading. First listen, then mark the right picture with your pencil. Here is the story.

Any child who is not thoroughly confused at this point has excellent listening comprehension! If, as is likely, the child erases all of it, he can do the test by following the example, but it places an extra premium on intelligence, and also upon personality--those easily discouraged probably score lower than they would if the instructions were brief and to the point. No child of this age could possibly retain all those items in memory, and most of them are irrelevant in any event. Test-taking skills will pay off even at this age! The Peabody Picture Vocabulary Test (1965) also has unnecessarily difficult instructions for children under 8 years: The addition of the following basic instructions not only adds to the memory load but also is conceptually

difficult for young children because it involves future projections, a conditional statement, and the desires of the examiner.

> When we get along further in the book you may not be sure you know the word, but I want you to look carefully at all the pictures anyway and choose the one you think is right.

Since this is an individual test, it is up to the examiner to see that the child knows how to respond, to encourage him to try, etc. The addition of a large comprehension problem and memory load can only detract from the real validity of the test. Primary Reading Profiles (SRA, 1962), Level 1 (end of first grade) has the following set of instructions for Test 4: Word Attack.

> Turn to this page, page 9, in your booklets. There are some little stories on this page and groups of pictures that go with the stories on this page. In each little story there is one word that you may not know. Read to yourself the story at the top of the first column. (Pause) You may not know the last word. What sound does it begin with? Yes, with the sound of t, as in take and Tom. Now look at the pictures that go with this story: a tie, a boat, a top, and a car. One of these pictures tells what the last word is. Could it be tie? No. The word tie begins with the sound of t, but it is not something to play with. Could it be boat? No. The word boat does not begin with the sound of t. Is it top? Yes, the word top begins with the sound of t, and a top is something to play with. So put an X in the circle below the picture of the top to show that you know the last word in the story is top. . .
>
> Now you know what to do with each little story. First you read the story to yourself. Then you make an X in the circle below the picture that tells what the last word is. If you do not know the last word in the story, think what sound it begins with and use the pictures to help you decide what the word is. Remember that the word you choose must make sense in the story. When you finish this page, go on and do the next two pages, also. The arrows at the bottom of the pages tell you to go on. The STOP sign will tell you where to stop. You will have 18 minutes for this part of the test. Ready, go!

In addition, for each item, there are three to four lines to be read just to find out whether the child can identify the last word as one of four pictures--a task that most tests can accomplish by use of only a single word!

SRA's Short Test of Educational Ability (SRA, 1966) also suffers from long instructions at a high verbal level. The Stanford Achievement Test (1964), Primary I (grades 1A and 2B), Test 6, Part C: Number Concepts, has three sets of instructions, given (and to be remembered) in sequence, already too large a memory load, but then an example is given from the third set, and children are told to go back to the first set, some further instructions are given for the first set, and then they are to do all three sets. It is interesting to note that the word problems at the Intermediate I Level have less verbal difficulty than the ones at the Primary I Level; it will be remembered that verbal difficulty adds to the memory load.

Ignoring of the memory load of instructions, however, is not limited to test writers; experimental child psychologists can be just as guilty, viz., Beilin's instructions to kindergartners giving a verbal rule for conservation (51 words, conditional sentences, and at least 8 semantic units) and a conservation of number rule (64 words and 9 semantic units).

Individual Test Items

The commonest way of introducing a memory screen into a test is by way of a question which requires memory for more units than can normally be retained. In some tests such questions occur only occasionally; in

others they are the rule. In either case there seems to be a lack of understanding of the limitations of children's memory store and even a single invalid item may seriously reduce the efficiency of these tests (which typically contain too few items for reliable measurement anyway), most particularly if the item tests a different set of abilities from those tested by the other items. An example of a test which overloads the short-term memory capacity on every item is Diagnostic Reading Tests (1966), Booklet I, Test IV: Story Reading (grade 1). This test, ostensibly of comprehension, has a very high memory load; the first item presents 12 separate, isolated facts to be retained by a child whose capacity is no more than three. And to add insult to injury, both delay and interference are introduced by interspersing the three comprehension questions among five questions about the structural properties of the English language in general. Further, there are only four independent items in the test, all of the same general format. A similar test for grade 2 contains an even heavier memory load; 25 items in the first "story." These are stories in name only; they resemble a string of almost randomly selected sentences, without any structure to the whole which might enable the reader to condense the isolated units to a small number of more inclusive semantic units, as normally occurs in coherent paragraphs for mature readers, but seldom in "simple" paragraphs written for beginners. The Durrell Analysis of Reading Difficulty (1955) likewise overtaxes the memory capacity of first graders. In Oral Reading, question 1 has a 4-unit load and question 2, a 9-unit load.

In Listening Comprehension, one item requires storage of 8 units of information, upon which 7 questions are asked, and here again memory for the sequential order of items is sometimes demanded. As we have seen, storage of order information takes up part of STM capacity. Their Reading Achievement, Test 2: Paragraph Comprehension is similarly constructed. The Gates-MacGinitie Reading Test (1965), Primary C (Speed and Accuracy for grades 2 and 3), is an example of a test of comprehension that does not overload memory; the paragraphs are coherent semantic wholes; the reader need not store an array of isolated facts; there is only one question on a paragraph; and the child must read the paragraph in order to answer correctly. There are other difficulties with this test, but it does show the possibility of constructing a reading test that is not largely a test of memory, at least no more so than ordinary comprehension of language demands. Primary Reading Profiles (1967) (grade 1A) Test 1, Aptitude for Reading, item no. 4, requires storage of at least 14 units, plus sequence, and inferences must be made from the information. Work done in short-term memory (inference, transformation, etc.) adds to the already impossible load. In the Pupil Progress Series (1956), Diagnostic Reading, Primary, Test 6: Reading to Locate Information, the memory load is so high that the child must search the paragraph for the answers to the questions. It is very similar to many adult tests of comprehension; it puts a premium on both STM and on perceptual speed in word recognition--skills which are of importance much later, but not at this early stage in reading.

The SRA Achievement Series (1958), Reading 1-2, contains 5 stories, which tend to be "laundry lists," each followed by 5 to 10 (!) questions which involve both facts and inferences. The Arithmetic 1-2 Test, Test 3: Number Problems, has a high memory load largely because of irrelevant statements which have nothing to do with the arithmetic but do add to the load: "Sally went to the store. She bought. . .etc." since the items are written and have pictures associated with them, the memory load may be lessened somewhat (if the child knows how to use the cues), but item No. 1 requires that 4 units be stored in memory and item No. 2, 6 units plus 3 conjunctions. It is not necessary to use more than 2 memory units to test the arithmetical skills at this level, as evidenced by the Comprehensive Tests of Basic Skills, or the Contemporary Math Test. The Stanford Achievement Tests (1964), Primary I, Test 3: Vocabulary, is intended to test general knowledge, but the procedure is such that the child must remember the whole statement (the examiner calls it a question, thereby causing confusion for literal-minded children) and the three possible choices--unless they can read (which is not being tested) or have phonic skills (not being tested either). To make matters worse, the lexical level is too high (not related to the general knowledge being tested), adding to the memory load; an example is the use of the word "enthusiastic" in a statement at the first grade level. Their Test 5: Word Study Skills, also makes a large memory demand (unless the child can read, which is not being tested). In this test, the examiner reads three words, then gives

some instructions, then says the key word, then reads the three words again--all this to test phonetic skills. There are good tests of phonetic skills that do not confound them with memory; such tests present two words and the child makes a simple "yes" - "no" judgment about their similarity. Since many more independent items can be given in the same period, such a test will be more reliable in any event. Test 6 - Part B: Problem Solving, contains oral arithmetic word problems which contain 3 or 4 units to be remembered. Some tests have fewer, as pointed out above, and so can test for arithmetical skills without an initial memory screening. In some items of this test irrelevant sentences are introduced, as in item No. 20. The first two sentences can be deleted without changing the problem in any way. Since children are not experts in knowing when an adult is just talking and when he is saying something, they are likely to try to store all the information, with the resulting overload and probable failure of the item. The Illinois Test of Psycholinguistic Abilties (1968) (Age 2-10), Test 2: Visual Reception (Visual Decoding), introduces a gratuitous memory factor. Here the child is shown a photograph of an object (a dog, for example). This is then covered, so the item must be stored in memory in some form. Then the child is shown a set of 4 photographs and is to pick out one "like" the first. But the correct choice is not identical to the original, it is merely from the same category--a quite different kind of dog--so that the cue is not visual similarity but categorial similarity, with a memory factor thrown in. The Sound Blending Test requires the memory for as many as 7 sounds (alleged to be phonemes),

to be blended together to produce a word. It is literally impossible for a human speaker to produce the proper allophone in isolation; it would have to be done by splitting a tape containing the word spoken normally and inserting silences between the phonemes. But the test does not provide a tape and demands instead that the examiners learn to do what cannot be done with human articulatory equipment. Even if the allophones were accurate, the child could not possibly remember so many; what he obviously does is to guess, given the first and last sounds and a general impression of length--and the guess will be the most frequent word which fits the characteristics (and that will, in fact, be correct). It is really a form of auditory Cloze test (cf. Peisach, 1965). However, the same strategy will not work as well for Part C of the test, where nonsense words consisting of 3 to 6 phonemes each are used. Here the contribution of memory to the test score will necessarily be even larger. There is no problem in building a sound blending test without a memory load, since one need only present two sounds, which should, of course, be the proper allophones, to be blended, use all the frequent diphones of English arranged in order of frequency (cf. Carterette & Jones, 1970), and avoid using real words which can be guessed. There is even a bonus in this arrangement, since many more independent items can be given in the same period of time, making for greater reliability. The Gates McKillop Reading Diagnostic Test (1962) uses the same technique, and has only a slightly smaller memory load.

In the Cooperative Tests (1965) - Primary Listening Test (grades 1, 2, and 3) - Form 23B, item No. 23, presumably a comprehension item, recites 6 items, not including the actor's name, telling what he was doing, and asks the child to pick out a picture showing what was missing from the list of six acts. Item No. 39 requires memory for four items in sequential order. The Reading Test from the same series (Form 23A) contains a story having 34 lines and many items of information. More information would be gained about comprehension if several short paragraphs were used with a question on each. The IPAT Culture-Fair Test (1950), Scale 1 (ages 4-8), seriously overloads memory, particularly at the younger level: The symbol substitution requires memory for six items--true they could be repeatedly checked, but not usefully in a speed test. Test 4, a vocabulary test, requires storage of four items, as does Test 7: Riddles, neither of which are intended to be memory tests. Test 5 is a test for following directions, but many questions require storage of four items plus sequential order and sometimes with conditional statements thrown in. The ETS Preschool Inventory (1967) (ages 3-6) has a four-unit memory load in item No. 19. Perhaps the most unexpected finding is that the Lorge-Thorndike Intelligence Tests for Level 1-2 contains no measure of memory span of any kind! The WISC does contain one such test (a rather poor digit span test) and avoids memory load in other subtests, although heaven knows what an imaginative child is to make of E's statement, "Here is a picture of a dog that has been cut up."

Temporal Relations

The third point to be illustrated is the memory load added by the temporal sequencing of the questions. Two examples here will suffice. In the Diagnostic Reading Tests (1966), Booklet I (grade 1) Test IV: Story Reading, the comprehension questions have been scattered among questions about the structure of English, so that the test is really testing delayed recall with verbal interference under the guise of comprehension. The same format is used for grades 2 and 3-4. The Pupil Progress Series (1956): Diagnostic Reading Test - Primary (grades 1 and 2B), Test 4, involves reading four stories in sequence, for speed, and then answering questions on all of them, the questions referring merely to the first, second, third, and fourth stories (this for first graders!) and calling for detailed factual answers. Not only is the memory load much too large, but again both delay and interference are added, and then the problem of reference to the proper original story serves to add to the confusion.

Conclusion

Sufficient evidence has been presented to show that many tests for young school children place an unintended load on short-term memory capacity. The argument is not that memory processes are unimportant but rather that they are such an essential ingredient of language comprehension and cognitive performance in general that they should be measured properly and accurately and not confounded with factors such as lexical or syntactic difficulty, language comprehension in general, arithmetical

skills, or concepts. Although the comments about memory factors in tests are based upon judgment rather than upon empirical tests, some support is given to the judgments by the fact that of three tests of general ability used by Feldman (1970) which did not prove to have any loadings on the memory factors in his battery, none were judged in this survey to have a memory loading. The tests were the Lorge-Thorndike Test I, the Metropolitan Word Meaning Test, and the Ammons Quick Test. Baumeister and Bartlett (1962) also report that STM is an important factor in many paper and pencil tests.

Since there appear to be large individual differences in memory span, both for unit storage and for sequence, and since these do not appear to be highly correlated with other unique skills involved in reading, listening, arithmetic and "general intelligence" (cf. Graham, 1964; Feldman, 1970), it seems wiser to insist that the memory skills be measured separately, and that the other relevant skills be assessed independently of the memory screen. This memory screen would appear to be much too pervasive in the testing of young children, particularly in the testing of language skills, and may, in part, account for the difficulties currently being experienced in the assessment of reading and in the diagnosis of reading difficulties.

REFERENCES

Aaronson, D., & Sternberg, S. Effects of presentation rates and signal-to-noise ratio on immediate recall. Paper presented at E.P.A., Philadelphia, 1964.

Atkinson, R.C., & Shiffrin, R.M. Human memory: A proposed system and its control process. Institute for Mathematical Studies in the Social Sciences, Stanford University, Technical Report No. 110, 1967.

Baumeister, A., & Bartlett, C.J. Further factorial investigations of WISC performances of mental defectives. American Journal of Mental Deficiency, 1962, 67, 257-261.

Beilin, H. Learning and operational convergence in logical thought and development. Journal of Experimental Child Psychology, 1965, 2, 317-339.

Bradley, P.A., Hoepfner, R., & Guilford, J.P. A factor-analysis of figural-memory abilities. Report, Psychological Labo - ratories, University of Southern California (Los Angeles), June, 1969, No. 43.

Broadbent, D.E. Techniques in the study of short-term memory. Acta Psychologica, 1965, 24, 220-233.

Calfee, R.C. Short-term retention in normal and retarded children as a function of memory load and list structure. Technical Report No. 75, Madison, Wisconsin: University of Wisconsin, Wisconsin Research and Development Center for Cognitive Learning, February, 1969.

Calfee, R.C., Hetherington, E.M., & Waltzer, P. Short-term memory in children as a function of display size. Psychonomic Science. 1966, 4, 153-154.

Carterette, E.C., & Jones, M.H. Alphabetic and phonemic texts of informal speech with statistical analyses. Berkeley, University of California Press (in press).

Feldman, B. Prediction of first grade reading achievement from selected structure-of-intellect factors. Doctoral Dissertation, University of Southern California, 1970.

Gage, N.L. Handbook of research on teaching. Chicago: Rand McNally, 1963.

Gordon, M.C., Gordon, A., & Perrier, C. Stimulus presentation rate and level of intellectual competency. Psychonomic Science, 1967, 8, 403-404.

References (Cont'd)

Graham, N.C. Short-term memory and syntactic structure in educationally sub-normal children. Language and Speech, 1964, 11 (4), 209-219.

Guilford, J.P. The nature of intelligence. New York: McGraw-Hill, 1967.

Haith, M. Information-processing in children. Lecture presented at University of California, Los Angeles, October 25, 1968.

Hansen, D.N. Short-term memory and presentation rates with young children. Psychonomic Science, 1965, 3, 253-254.

Hurlock, E.B., & Newmark, E.D. The memory span of pre-school children. Journal of Genetic Psychology, 1931, 39, 157-173.

Inglis, J., & Sykes, D.H. Some sources of variation in dichotic listening performance in children. Journal of Experimental Child Psychology, 1967, 5, 480-488.

Jones, M.H. The effect of linguistic transformations in short-term memory in children. Western Psychological Association Annual Meeting, 1968.

Jones, M.H. Children's short-term recognition memory for item and for sequence for two kinds of material. Paper presented at Western Psychological Association, Los Angeles, 1970.

Kennedy, G.D. Children's comprehension of English sentences comparing quantities of discrete objects, Doctoral Dissertation, University of California, Los Angeles, 1970 a.

Kennedy, G.D. The language of tests for young children. Los Angeles: University of California, Graduate School of Education, Center for the Study of Evaluation, Working Paper No. 7, February, 1970 b.

Lloyd, K.E. Supplementary report: Short-term retention as a function of average storage load. Journal of Experimental Psychology, 1961, 62, 632.

Mackworth, J.F. Interference and decay in very short-term memory. Journal of Verbal Learning and Verbal Behavior, 1964, 3, 300-308.

Mathews, W.A. Transformational complexity and short-term recall. Language and Speech, 1968, 11, 120-128.

References (cont'd)

Meeker, M. Immediate memory and its correlates with school achievement. Doctoral Dissertation, University of Southern California, 1966.

Moore, M.E., & Ross, B.M. Context effects in running memory. Psychological Reports, 1963, 12, 451-465.

Mortenson, F.J., & Loess, H. Effect of very brief interpolated activity on short-term retention. Perceptual and Motor Skills, 1964, 18, 797-803.

Murdock, B.B., Jr., & Walker, K.D. Modality effects in free recall. Journal of Verbal Learning and Verbal Behavior, 1969, 8, 665-676.

Norman, D.A. Memory and attention: An introduction to human information processing. New York: Wiley, 1969.

Orpet, R.E., & Meyers, C.E. Six structure-of-intellect hypotheses in six-year-old children. Journal of Educational Psychology, 1966, 57, 341-346.

Peisach, E.C. Children's comprehension of teacher and peer speech. Child Development, 1965, 36, 467-480.

Phillips, J.L., Shiffrin, R.M., & Atkinson, R.C. Effects of list length on short-term memory. Journal of Verbal Learning and Verbal Behavior, 1967, 6, 303-311.

Posner, M.I., & Rossman, E. Effect of size and location of information transforms upon short-term retention. Journal of Experimental Psychology, 1965, 70, 496-505.

Ross, B.M. Sequential visual memory and the limited magic of the number seven. Journal of Experimental Psychology, 1969, 80, 339-347.

Schulman, A.I., & Lovelace, E.A. Recognition memory for words presented at a slow or rapid rate. Psychonomic Society Meeting, 1967 (abstract).

Sitterly, T.E. Short-term retention of sequentially presented digits as a function of interdigit interval, digit duration, and series length. Journal of Experimental Psychology, 1968, 78, 174-178.

References (cont'd)

Sperling, G., & Speelman, R.G. The effect of sound stimuli on short-term memory. Bell Telephone Laboratories Technical Memorandum, August, 1967.

Talland, G.A. Short-term memory with interpolated activity. Journal of Verbal Learning and Verbal Behavior, 1967, 6, 144-150.

Tyler, L.E. The psychology of human differences. New York: Appleton-Century-Crofts, 1965.

Whimbey, A., Fischhof, V., & Silikowitz, R. Memory span: A forgotten capacity. Journal of Educational Psychology, 1969, 60, 56-58.

Whimbey, A.E., & Leiblum, S.L. Individual differences in memory span with and without activity intervening between presentation and recall. Journal of Educational Psychology, 1967, 58, 311-314.

TESTS CITED

Comprehensive Tests of Basic Skills. Arithmetic. Level 1 (grades 2, 5, 3 and 4). California Test Bureau, 1968.

Contemporary Mathematics Test. California Test Bureau. Lower Elementary, Upper Elementary Forms, 1965.

Cooperative Tests. Primary. Reading. (grades 1-3). Educational Testing Service, 1965.

Diagnostic Reading Tests. Committee on Diagnostic Reading Tests, Inc., Mountain Home, North Carolina, 1966.

Durrell Analysis of Reading Difficulty, New Edition. Harcourt, Brace and World, 1955.

Gates-MacGinitie Reading Test (Kindergarten-Grade 12). Teachers College Press, 1962.

Gates-McKillop Reading Diagnostic Tests. Form 1, (grades 1-12). Teachers College Press, 1962.

Illinois Test of Psycholinguistic Abilities. University of Illinois Press, 1968.

IPAT Culture Fair Intelligence Test. Institute for Personality and Ability Testing, 1950.

Peabody Picture Vocabulary Test, American Guidance Service, 1965.

Pre-school Inventory. Educational Testing Service, 1967.

Primary Mental Abilities. Science Research Associates, 1962. Rev.

Primary Reading Profiles. Houghton Mifflin, 1967 edition.

Pupil Progress Series. Diagnostic Reading. Primary, Form A., Level I (grades 1, and 2B). Scholastic Testing Service, 1956.

Short Test of Educational Ability. Science Research Associates, 1966.

SRA Achievement Series. Reading 1-2, Form C. Science Research Associates, 1958.

Stanford Achievement Test. Harcourt, Brace, and World, 1964.

Wechsler, D. Wechsler Intelligence Scale for Children. Psychological Corporation, 1949.

This publication is published pursuant to a contract with the U.S. Office of Education, Department of Health, Education and Welfare. Points of view or opinions stated do not necessarily represent official U.S.O.E. position or policy.